

WO0142921

Publication Title:

MULTIPLE COPY CAPABILITY FOR NETWORK BACKUP SYSTEMS

Abstract:

A method and apparatus for providing a multiple copy file backup system (210) is disclosed in which files that have been previously backed up by a first user are to be backed up by different users. The first instance of backing up a file results in storing the file and user in an entry in 238808-1a file data table (212). For other users needing to backup the same file, the contents of the file are copied from the first location into a second location within the file backup storage system (210). The file backup system creates two new entries in the file data table (212). In one entry the file stored in the second location is associated with a User-Pool that allows sharing the file contents among the users. The second and subsequent entries in the file data table (212) will be created as the subsequent users desiring to backup the file. These entries will indicate that the file is provided by the User-Pool and the necessary data will be copied from the User-Pool entry to properly associate the user with the file contents.

Data supplied from the esp@cenet database - <http://ep.espacenet.com>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 June 2001 (14.06.2001)

PCT

(10) International Publication Number
WO 01/42921 A1

- (51) International Patent Classification⁷: **G06F 11/00**
- (21) International Application Number: **PCT/US00/42801**
- (22) International Filing Date:
13 December 2000 (13.12.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/170,341 13 December 1999 (13.12.1999) US
- (71) Applicant: **CONNECTED CORPORATION [US/US];**
24 Prime Parkway, Natick, MA 01760 (US).
- (72) Inventors: **CANE, David, A.;** 32 Kinnaird Street, Cambridge, MA 02139 (US). **VAITZBLIT, Lev;** 283 Old Pickard Road, Concord, MA 01742 (US). **FREIDIN, Jonathan, F.;** 23 Pine Cliff Drive, Marblehead, MA 01945 (US). **SPEARE, Philip, A.;** 118 Ridge Street, Arlington, MA 02174 (US).
- (74) Agents: **SCHURGIN, Stanley, M. et al.;** Weingarten, Schurgin, Gagnebin & Hayes LLP, 10 Post Office Square, Boston, MA 02109 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: **MULTIPLE COPY CAPABILITY FOR NETWORK BACKUP SYSTEMS**

(57) Abstract: A method and apparatus for providing a multiple copy file backup system is disclosed in which files that have been previously backed up by a first user are to be backed up by different users. The first instance of backing up a file results in storing the file in a first location and associating the file and user in an entry in 238808-1a file data table. For other users needing to backup the same file, the contents of the file are copied from the first location into a second location within the file backup storage system. The file backup system creates two new entries in the file data table. In one entry the file stored in the second location is associated with a User-Pool that allows sharing the file contents among the users. The second and subsequent entries in the file data table will be created as subsequent users desiring to backup the file. These entries will indicate that the file is provided by the User-Pool and the necessary data will be copied from the User-Pool entry to properly associate the user with the file contents.

WO 01/42921 A1

-1-

TITLE OF THE INVENTION
MULTIPLE COPY CAPABILITY FOR NETWORK BACKUP SYSTEMS

5 CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority under 35 U.S.C. §119(e) to provisional patent application serial no. 60/170,341 filed December 13, 1999; the disclosure of which is incorporated by reference.

10

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR
DEVELOPMENT

N/A

15

BACKGROUND OF THE INVENTION

This application relates generally to file backup systems, and more particularly to file backup systems in a client-server system having a multiple copy capability.

20 Client-server network systems are well known and widely used in many industries and applications. In a typical client-server system, users operating a client machine send data to one or more central computers—the server—for processing. The processed data may be stored
25 centrally on the server or returned to the client for local storage. In either case, a failure of the primary data storage system whether on the client or the server, can result in a catastrophic loss of the user's data. To prevent this loss of data due to the failure of the
30 primary data storage system, a file backup system is commonly used to allow the recovery of the client data.

The backup system for a client-server system typically maintains an extra copy of the user's data on the server. The backup system typically maintains this extra copy of the user's data on another storage device
5 such as a serial tape device that is less expensive than the hard drives used on the client and server systems.

In a client-server system more than one user may attempt to backup a particular data file. If left unchecked, the backup system could be overwhelmed by the
10 repeated backup of a single data file. Previously, backup systems have been developed that require only the first user to transmit a specific data file to the server for storage on the backup system. These systems detect whenever a subsequent user attempts to backup an
15 identical data file and do not require the second user to transmit the data file and do not store the user's data file repeated times.

A problem arises however, if a user has been on the system for a period of months or years and now has
20 references to shared backup data files that may located in many physical locations throughout the system. Some physical storage systems, such as a tape system, allow only sequential access to the data contained therein. As discussed above, backup systems that support a large
25 number of users typically migrate backup data from their high-speed, random-access disk storage systems to an automated tape library (ATL) located at the server. An ATL provides a very low cost per unit of storage and may also require less floor space, also reducing cost. ATL's
30 do, however, have a longer time associated with recovering data from a tape. In particular, while the

data for a particular user may be logically grouped so that it is only on one tape, the references to shared files from other users may be contained on many other physically separate tapes. Each tape would need to be
5 separately read from to retrieve the files and the data contained therein. In addition, an ATL system can be considered unreliable since the probability of losing a tape over the life time of the system is high.

Additionally, if a user's account is cancelled, the
10 present backup systems typically do not allow the space used by the cancelled account to be reclaimed, since removing the files of the cancelled account may result in the loss of shared data. To avoid the indiscriminate loss of shared data, each file must be checked to see if
15 it is shared among other users in what is a time consuming and laborious task.

It would therefore be desirable to be able to back up files in such a way that only a small number of tapes need to be read and that the data can be easily retrieved
20 from. It would also be desirable to allow a file backup system retrieve storage space when an account is cancelled in a more time efficient manner.

BRIEF SUMMARY OF THE INVENTION

25 A method and apparatus for providing a multiple copy file backup system is disclosed in which files that have been previously backed up by a first user are to be backed up by different users. The first instance of backing up a file results in storing the file in a first
30 location and associating the file and user in an entry in a file data table. For other users needing to backup the

same file, the contents of the file are copied from the first location into a second location within the file backup storage system. The file backup system creates two new entries in the file data table. In one entry the
5 file stored in the second location is associated with a User-Pool that allows sharing the file contents among the users. The second and subsequent entries in the file data table will be created as subsequent users desiring desire to backup the file. These entries will indicate
10 that the file is provided by the User-Pool and the necessary data will be copied from the User-Pool entry to properly associate the user with the file contents.

A method and apparatus for providing a multiple copy file backup system is disclosed in which files that are
15 to be backed up by different users are copied into a second location within the file backup storage system and associated with a User-Pool that allows sharing the file contents among the users.

In one embodiment a file data table is created by a
20 backup file system coupled to a server in a client-server system. A user of a client machine coupled to the server requests to backup a file, the server searches the file data table for any entries that contain as the File ID field the file identifier provided by the user. If no
25 entry is found, the backup file system requests the contents of the file from the client machine and stores the contents of the file at a first location. A new entry in the file data table is created and a User-ID field in the new entry in the file data table is set to
30 the user identifier of the user, a File-ID field in the new entry in the file data table is set to the file

identifier, a Sourced By field in the new entry in the file data table is set to the user identifier, a File Location pointer field in the new entry in the file data table is set to the first location, and a Share Count
5 field in the new entry in the file data table is set to zero.

If one entry in the file data table is found, the contents of the file at the first location are copied to a second location and two new entries in the file data
10 table are created. In one of the two new entries that associates the newly copied file with a User-Pool identifier, the User-ID field in the new entry in the file data table is set to the user identifier of the User-Pool, the File-ID field in the new entry in the file
15 data table is set to the file identifier, the Sourced By field in the new entry in the file data table is set to the User-Pool, the File Location pointer field in the new entry in the file data table is set to the second location, and the Share Count field in the new entry in
20 the file data table is set to one.

As discussed above, the server creates two new entries in the file data table. The second new entry created by the server will include the User-ID field in the new entry in the file data table is set to the user
25 identifier of the user, the File-ID field in the new entry in the file data table is set to the file identifier, the Sourced By field in the new entry in the file data table is set to the User-Pool, the File Location pointer field in the new entry in the file data
30 table is set to the second location, and the Share Count

field in the new entry in the file data table is set to zero.

5 If two or more entries in the file data table are located, the entry associating the file identifier with the User-Pool is located from among the entries. The server creates a new entry in the file data table. In the new entry the User-ID field in the new entry in the file data table is set to the user identifier of the user, the File-ID field in the new entry in the file data
10 table is set to the file identifier, the Sourced By field in the new entry in the file data table is set to the User-Pool, the File Location pointer field in the new entry in the file data table is set to the second location as read from the User-Pool entry associated with
15 the file identifier, and the Share Count field in the new entry in the file data table is set to zero. The Share Count field in the entry associating the file identifier with the User-Pool is incremented by one.

20 BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

The invention will be more fully understood by reference to the following Detailed Description of the Invention in conjunction with the Drawing of which:

25 Figs. 1A-1D are a flow chart illustrating a method of storing more than one copy of a file on a file backup system in accordance with the present invention;

Fig. 2 is a block diagram of the present invention;

Fig. 3 is a schematic representation of a file data table in accordance with the present invention; and

Fig. 4 is a schematic representation of the file data table and the file backup storage device in accordance with the present invention.

5

DETAILED DESCRIPTION OF THE INVENTION

A method and system in accordance with the present invention for maintaining multiple backup copies on a backup system is disclosed. Typically, the present method and apparatus are part of a client-server system in which users operating a client machine send and receive data from a server. The server maintains backup copies of the user's data on a storage system which may be contained within the server or contained within a separate backup system coupled to the server.

15

Figs. 2 and 3 illustrates a system operative to perform multiple copy backup file storage in accordance with the present invention. A plurality of users, User-A 202, User-B 204, to User-N 206 operate a plurality of client machines 202A, 202B and 202N respectively. The client machines 202A, 202B and 202N are coupled to a server 206 that provides data processing and at least some data storage. A backup system 210 that comprises a file data table 212 and a file backup storage system 214 is illustrated as being coupled to the server 206. However, the backup system 210, or a part of the backup system 210 such as the file data table 212, may also be included within the server 204 as well. For example, the file data table 212 may be stored on the storage system within the server and the file backup storage system may be a separate storage system or a storage system on another server (not shown) or a client machine(not

30

shown). The file backup storage system 214 may be a sequential accessible tape or other automated tape library (ATL) system, a random access hard-drive, read/writeable CD, a non-volatile electronic memory such as a Flash-RAM or other non-volatile RAM, or any other storage media that allows the storage and accessing of data.

Fig. 3 illustrates one embodiment of the File Data Table 212. The File Data Table can include a plurality of data associated with each file that is stored in the file backup storage system by the user. In the illustrated embodiment, the file data table 300 includes a User-ID field 302 indicative of the user storing the file in the backup system and a File-ID field 304 of the file being stored. Both the User-ID and the File ID fields should contain unique user and file identifiers respectively to prevent confusion and a loss of data. In addition, the file data table 300 can include a File Location pointer field 306 that points to the location of the contents of the stored file within the backup system, which may be a logical pointer or a physical pointer depending on the type of system used. The file data table further includes a Sourced By field 308 that includes the user identifier of the user providing the copy of the file. The file data table 300 can also contain a Share Count field 310 that indicates the number of users that may share the file copy and a other data considered pertinent by the user may be stored in an Other data field 312. It should be understood that the file data table 212 could be any form of a data structure that would allow data to be stored and accessed. For

example, the file data table may be, without limitation, an array of data structures, a matrix, or a linked list.

Referring again to figure 2, the server 206, using the file data table 212, maintains the associations
5 between particular users and the files backed up by those users. This is illustrated schematically in Fig. 2 in that the file backup storage system 214 is divided into a plurality of areas corresponding to the plurality of users. For example, area 216 contains the files backed
10 up by User-A, area 218 contains the files backed up by User-B, and area 220 contains the files backed up by User-N. The groupings of the files within these areas is for schematic purposes only, and does not imply that the files are actually located within the same physical area
15 of the file backup storage system 214. Rather the files are logically associated with each user through the file data table 212.

In the first instance of each file being backed up by a first user, the contents of the file will be stored
20 at a first location and a new entry created in the file data table 212. The server will set the user-ID field and the Sourced By field of the file data table 212 to the unique user identifier of the user performing the operation, and the server will set the File ID field will
25 to the unique file identifier associated with the file to be backed up. In addition, the File Location pointer field will be set to the first location, which may be a physical location or a logical association within the storage system, at which the contents of the file are
30 stored, and the Share Count field will be set to zero (0) indicating that no other files share this data.

In the event that another user requests to make a backup copy of a file previously backed up by a first user, the server will make a second copy of the file contents and store the second copy at a second location within the file backup storage system 214. The server associates the second copy of the file contents with a User-Pool of files. The User-Pool contains the second copy of any file that has been backed up more than once by more than one user. Other users that wish to back up a copy of a file, that has been previously backed up, may access the file within the User-Pool.

For each file stored in the file backup storage system and associated with the User-Pool there will be at least two entries created in the file data table 212. One entry, created when the file is placed in the User-Pool, will set the User-ID field to the unique User-Pool identifier, the File-ID field to the unique file identifier, and the Sourced By field to the unique User-Pool identifier. In addition, the server will set the File Location pointer field to point to the file contents stored at the second location within the file backup storage system 214, and the Share Count field will be set to one (1) for the second user needing to backup the file and incremented for each subsequent user thereafter.

The second entry in the file data table 212 will associate the second user needing to backup the file with the previously backed up file. In this entry, the server will set the User-ID field to the unique user identifier of the user requesting to back up the file, set the File-ID field to the unique file identifier associated with the file to be backed up, and set the Sourced By field to

the unique User-Pool identifier. In addition, the server will set the File Location pointer field to point to the file contents stored at the second location within the file backup storage system 214, and the Share Count will
5 be set to zero (0). For each subsequent user that needs to backup the previously backed up file, a new entry will be created in the file data table 212 associating the user with the copy of the file stored in the User-Pool area as described above.

10 Depending on the availability of system resources, the server may select another time in the future to actually create and store the second copy of the file being backed up. In this case, the File Location pointer field in the User Pool entry in the file data table will
15 be set to point to the first location at which the file is stored. Each of the files to be backed up at a second location can be placed in a queue that is then accessed at the selected time. In addition, the order that the files in the queue are accessed may be manipulated to
20 allow for more efficient storage on the backup storage system. For example it may be more efficient to store certain files physically close to one another on a tape in an ATL.

In the event that a user's account is to be
25 cancelled and the files that the user backed up removed, the server will search the entries in the file data table 212 for each entry in which the User ID field and the Sourced By field are both set to the unique user identifier of the user whose account is to be cancelled.
30 The server then instructs the file backup storage system to delete the contents of the files pointed to by the

File Location pointer fields of the respective entries. In this way, each of the files that are unique to that user are removed, but files that are shared by other users are not removed. The server searches the entries
5 in the file data table 212 for any entry having a User ID filed set to the unique user identifier of the user whose account is being removed and the Sourced By field is set to the User Pool identifier and deletes those entries in the file data table. In this way, all entries associated
10 with the desired user are removed from the file data table 212 but entries associated with other users sharing the data are not. In addition, the server decrements the Share Count field in the User Pool entries in which the user being removed shared data. Once the Share Count
15 field has been decremented, the server may determine if the Share Count is now zero, and if so, the server may delete the file pointed to in the File Location pointer field, and the entry in the file data table.

As discussed above, the present invention may be
20 used with many different file backup systems and storage systems. As discussed above, a file tape back up system may not be considered reliable over the lifetime of the system because the probability of losing at least one tape is very high. Making a copy of the file data
25 contained in the User Pool will increase the robustness of the overall system by preventing a single point failure from destroying the only copy of the backed up data. This extra copy of the User Pool file data may be made and stored in a third location that is physically
30 separate from the first and second copies, such as a separate tape or another hard drive. This extra copy of

the User Pool data should be stored at a location that is the same logical location as the original User Pool data, but at a separate location or system, so that the File Location pointer fields in the User Pool entries in the file data table will point to the extra copy of the data when necessary.

In another embodiment, multiple tables may be used to store the file data and the User Pool data separately to facilitate maintenance of each table and the files contained within and also to save on storage space within each table. For example, the file data table in which each user is associated with the files backed up by that user will not contain a Share Count field. In addition, a Sourced By field may not be necessary in the file data table as well, since each user will source those files to itself.

The operation of the presently described multiple copy backup system is described below with respect to the flow diagrams illustrated in Figs. 1A-1D. As depicted in Fig. 1A, the server receives a unique user identifier of a user desiring to backup a the contents of a file identified by a unique file identifier also provided by the client machine, as illustrated in step 102. The server creates a first new entry in the file data table, as illustrated in step 104, and sets the User ID field to the unique user name provided by the client machine and the File ID field to the unique file ID also provided by the client machine, as illustrated in steps 104, 106, and 108 respectively. The server searches the file data table for the file identifier, as illustrated in step 110. If the file identifier is not located within the

file data table, control is passed to step 150, as illustrated in step 112. If the file identifier is located within the file data table, the server reports to the client that the file has been previously stored on the backup system, as illustrated in step 114. The server determines if there are two or more entries in the file data table having a File-ID field set to the unique file identifier received from the client machine, as illustrated in step 116. If there are not two or more entries in the file data table 212 having the File-ID field set to the unique file identifier received from the client machine, control passes to step 130. If there are two or more entries in the file data table 212 having the File-ID field set to the unique file identifier received from the client machine, control passes to step 120.

If one copy of the file exists on the backup system then the server copies the file currently stored on the backup system, as illustrated in step 130 and stores the new copy of the file contents at a second location within the file backup storage system 214, as illustrated in step 132. The server 206 sets the User-ID field identifier to the received unique User-Pool identifier, as illustrated in step 136 and further sets the File Location pointer field to point to the contents of the file stored at the second location, as illustrated in step 138. In addition, the server sets the Share Count field to one (1), the Sourced By field to the unique User-Pool identifier, and the File-ID field to the unique file identifier received from the client machine, as illustrated in steps 140, 142, and 144 respectively.

The server then sets the remaining fields in the first new entry in the file data table 212 previously created in step 104. The server sets the Sourced By field of the first new entry to the unique User-Pool identifier, the File Location pointer field is set to point to the contents of the file stored at the second location, and the Share Count field is set to zero (0). Control is then returned to step 102.

As discussed above, because it may not be the most efficient user of the processor and storage resources of the server to immediately store the second copy of the file, the server may place the file identifiers into a queue to be saved at a later time. Therefore, in another embodiment, when the User Pool entry is made in the file data table, initially the File Location pointer field will point to the first location of the file. The file identifier is placed into a queue and accessed at a later time for storage at the second location. Only after the file is saved in the second location will the File Location pointer field in the User Pool entry and the first new entry be set to the second location at which the file is saved.

If at step 118 there are two or more entries in the file data table in which the File-ID field is set to the unique file identifier received from the client, control passes to step 120. The server identifies the User-Pool entry in the file data table associated with the unique file identifier as illustrated in step 120. The server then sets the remaining fields in the first new entry in the file data table 212. The server sets Sourced By field to the unique User-Pool identifier, as illustrated

in step 122. The server then sets File Location pointer field to the value stored in the File Location pointer field of the User-Pool Identified field identified in step 120, as illustrated in step 124. In this way the new entry in the file data table will point to the file contents located at the second location that is associated with the User-Pool. The server then sets the Share Count field in the first new entry to zero (0), as illustrated in step 126. In addition, the server increments the Share Count field in the User-Pool entry identified in step 120, as illustrated in step 128. Control is then returned to step 102.

If at step 106, the received file identifier is not located within the file data table entries, control passes to step 150 as described above. As illustrated in step 150, the server requests the client machine to transmit the contents of the file to the server. The server receives the file contents, as illustrated in step 152, stores the file contents at a first location on the file backup storage system 214 as illustrated in step 154. The server sets the File Location pointer field of the first new entry in the file data table to the point to the file contents located at the first location, as illustrated in step 156. As illustrated in step 158, the server sets the Sourced By field to the unique user identifier received from the client in step 102. As illustrated in step 160, the server sets the share count field of the first new entry in the file data table to zero (0). Control is then returned to step 102.

An exemplary illustration of the relationship between the file data table and the stored files in the

backup system is illustrated in Fig. 4. An exemplary file data table 401 is provided in which the User-ID is provided in field 402, the File-ID is provided in field 404, the File Location pointer is provided in field 406, the Sourced By identifier is provided in field 408, and the Share Count is provided in field 410. An extra field 411 is provided for other data. An exemplary disk storage system 412 is provided in which the file contents are stored for various files at unique file server locations. In the illustrated examples the various data are purely exemplary and are not meant in any way to be limiting.

Entry 403 in the file data table 401 includes a User-ID AAA, a File-ID A1, a File Location pointer 100 and a Share Count of zero. Because this was the first instance of saving file A1, the Sourced By identifier field is set to the user identifier AAA. The corresponding entry in the disk storage system is entry 416 in which the contents of the file A1 are stored at file server location 100. Entry 405 in the file data table 401 and the corresponding entry 421 similarly identify the file B2 as belonging to user AAA with the contents of the file are stored at file server location 200.

Entry 407 in the file data table 401 includes a user identifier BBB that has also needed to backup file B2 a second time. In this instance, the contents of the file B2 have been copied to the backup storage system a second time and stored at location 1000. In this instance two entries have been made in the file data table. The first entry is entry 407 in which the User-ID field and the

File-ID field are set. The User-Pool entry is created and the data set to enable the proper data to be provided to entry 407.

5 The entry 411 is created that associates the file B2. The entry 411 includes the User-ID field set to the unique User-Pool identifier, the File-ID is set to the unique file identifier, the File Location pointer field is set to 1000, the second location at which the contents of the file are stored, the Sourced By field is set to
10 the unique User-Pool identifier, and the Share Count field is set to one (1). The corresponding entry in the backup storage system 412 is entry 425, located at file server location 100 and storing the contents of the file B2 in a second location. The entry 407 is then provided
15 with the correct data corresponding to the file B2. The File Location pointer field is set to 1000 and the Sourced By field is set to the unique User-Pool identifier.

Entry 409 in the file data table 401 includes a
20 User-ID field set to CCC and a File-ID field set to B2. This is the third instance of the file B2 being backed up by a distinct user on the backup storage system. The entry 409 is the fourth entry in the file data table 401 having a File-ID field set to the identifier B2. Thus,
25 the entry 409 includes the File Location pointer field set to 1000, the second location of the contents of the file B2, and the Sourced By field is set to the unique User-Pool identifier. In addition, the Share Count field of the User-Pool entry 411 associated with the second
30 copy of the file B2 is incremented again, to two (2) to

show that two users are currently sharing the contents of the file from the user pool.

In this way, any number of users may share the contents of the file B2, or any other file stored in the user pool area without requiring an increase in the number of times that the file B2 is stored on the system. The number of file data table entries will increase by one with each user that wishes to access a file stored in the user pool area. Each new entry will identify the particular user and the particular file and the particular file server location of the file.

As discussed above, many different types of storage media may be used to provide the file backup storage system. For some of these media, access costs and storage costs are such that the costs associated with backing up files makes it more economic to provide files to the User Pool after 3, 4, or even more files have been saved for individual users. It would be obvious to one of skill in the art, therefore, to modify the above described multiple copy file backup system by selecting a threshold number before a second copy of the file is created in the User-Pool to a number greater than two.

Those of ordinary skill in the art should further appreciate that variations to and modification of the above-described methods and systems for providing a multiple copy capability for a network backup system may be made without departing from the inventive concepts disclosed herein. Accordingly, the invention should be viewed as limited solely by the scope spirit of the appended claims.

CLAIMS

1. A method for providing multiple copies of backup files on a file backup system, the file backup system storing a plurality of files and a file data table containing a plurality of file data associated with each of the plurality of files, a server coupled to the file backup system and to a client machine, the client machine operated by a user having a user identifier, the method comprising the steps of:
- 5 providing by the client machine to the server a file identifier identifying a file to be stored on the file backup system
- providing by the client machine to the server the user identifier of the user operating the client machine;
- 15 creating a first new entry in the file data table;
- setting one of the plurality of file data of the first new entry in the file data table indicative of the file identification to the file identifier provided by the client machine;
- 20 setting one of the plurality of file data of the first new entry in the file data table indicative of the user identification to the user identifier provided by the client machine;
- detecting, by the server, if the file identified by the file identifier is presently stored on the file backup system;
- 25 in the event that a single copy of the file identified by the file identifier is presently stored on the file backup system;
- 30 copying the contents of the single copy of the file to a second location within the file backup system;

creating a second new entry in the file data table;

5 setting one of the plurality of file data of the second new entry in the file data table indicative of the file identification to the file identifier provided by the client machine;

10 setting one of the plurality of file data of the second new entry in the file data table indicative of the user identification to a user identifier indicative of a user pool;

 setting one of the plurality of file data of the second new entry in the file data table indicative of the source by identification to the user pool identifier;

15 setting one of the plurality of file data of the second new entry in the file data table indicative of the number of users sharing the file contents one;

20 setting one of the plurality of file data of the first new entry in the file data table indicative of the file location to point to the to the second copy of the file;

 setting one of the plurality of file data of the first new entry in the file data table indicative of the sourced by identification to the user pool identifier;

25 setting one of the plurality of file data of the first new entry in the file data table indicative of the file location to point to the to the second copy of the file.

30 2. The method of claim 1 further including the steps of:

in the event that the file associated with the file identifier is not presently stored on the backup system, receiving by the server the file contents from the client,

5 storing the file on the backup system, and
creating a new entry in the file data table
setting one of the plurality of file data of
the new entry in the file data table indicative of the
file identification to the file identifier provided by
10 the client machine;

setting one of the plurality of file data of
the first new entry in the file data table indicative of
the user identification to the user identifier provided
by the client machine;

15 setting one of the plurality of file data of
the first new entry in the file data table indicative of
the sourced by identification to the user identifier
provided by the client machine;

20 setting one of the plurality of file data of
the first new entry in the file data table indicative of
the number of users sharing the file contents to zero;

setting one of the plurality of file data of
the first new entry in the file data table indicative of
the number of users sharing the file contents to zero;

25 setting one of the plurality of file data of
the first new entry in the file data table indicative of
the file location to point to the to the second copy of
the file.

30 3. The method of claim 1 further including the steps
of:

in the event that two or more copies of the file associated with the file identifier are presently stored on the backup system,

5 creating a new entry within the file data table associated with the stored file;

 setting one of the plurality of file data of the new entry in the file data table indicative of the user identification to the user identifier provided by the client machine;

10 setting one of the plurality of file data of the new entry in the file data table indicative of the file identification to the file identifier provided by the client machine;

 setting one of the plurality of file data of the new entry in the file data table indicative of the source by identification to the user pool identifier;

 setting one of the plurality of file data of the new entry in the file data table indicative of the number of users sharing the file contents zero;

20 finding within the file data table the entry having a user identification and a sourced by identification set to the user pool identifier and having a file identification set to the file identifier received from the client machine;

25 incrementing the share count of the located file table entry

 setting one of the plurality of file data of the first new entry in the file data table indicative of the file location to point to the to the file location contained within the found file table entry.

30

4. The method of claim 1 further including the steps of:

receiving at the client a user identifier to be cancelled from the file backup system;

5 finding in the file data table, entries having a user identifier and a sourced by identifier set to the user identifier to be removed;

removing the file contents located at the location pointed to by the file location entry within each of the
10 found file data table entries.

5. The method of claim 1 wherein the step of copying the contents of the single copy of the file to a second location includes copying the contents of the single copy
15 of the file to a second location within the file backup system.

6. The method of claim 1 wherein the step of copying the contents of the single copy of the file to a second
20 location includes copying the contents of the single copy of the file to a new second location within a queue and storing the contents of the single copy of the file to a second location within the file backup system at a later time and setting one of the plurality of file data of the
25 first new entry in the file data table indicative of the file location to point to the to the new second copy of the file.

7. The method of claim 1 further comprising the steps
30 of copying the contents of the file stored at the second location to a third location.

8. A method for providing a multiple copies of backup files on a file backup system, the method comprising the steps of:

- 5 making a first copy of a file on the backup system by a first user;
 - creating a first entry in a file data table at a first location;
 - associating the first file with the first user as a
 - 10 file source within the first entry;
 - associating the first file with the first location within the first entry;
 - requesting to make a second copy of the file by a second user;
 - 15 making a second copy of the file in a second location;
 - creating a second entry in a file data table;
 - associating the second file with a user pool as a file owner and a file source within the second entry;
 - 20 associating the second file with the second location within the second entry;
 - creating a third entry in the file data table;
 - associating the second file with the second user as and identifying the source of the file as the user pool
 - 25 within the third entry; and
 - associating the second file with the second location within the third entry.

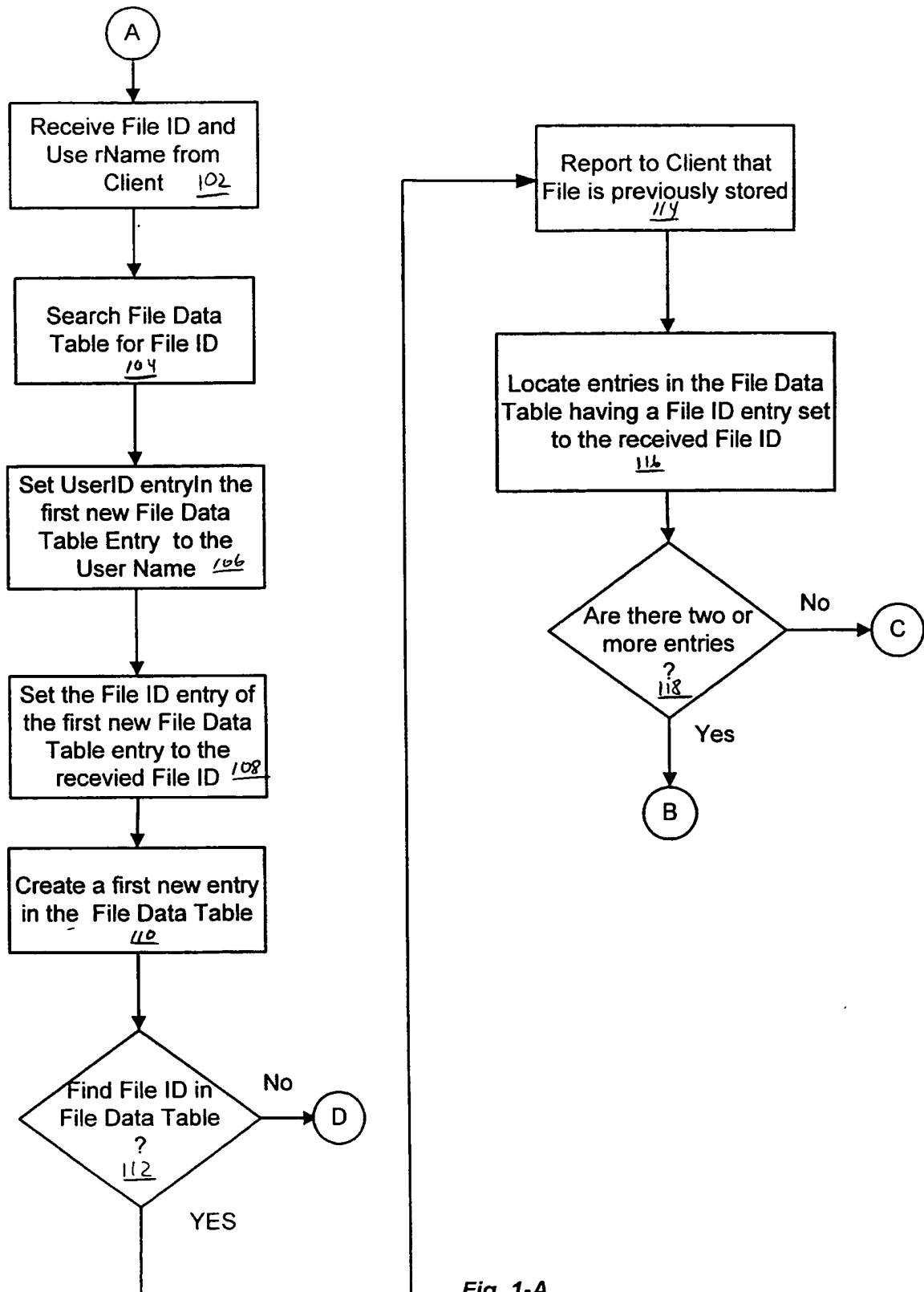


Fig. 1-A

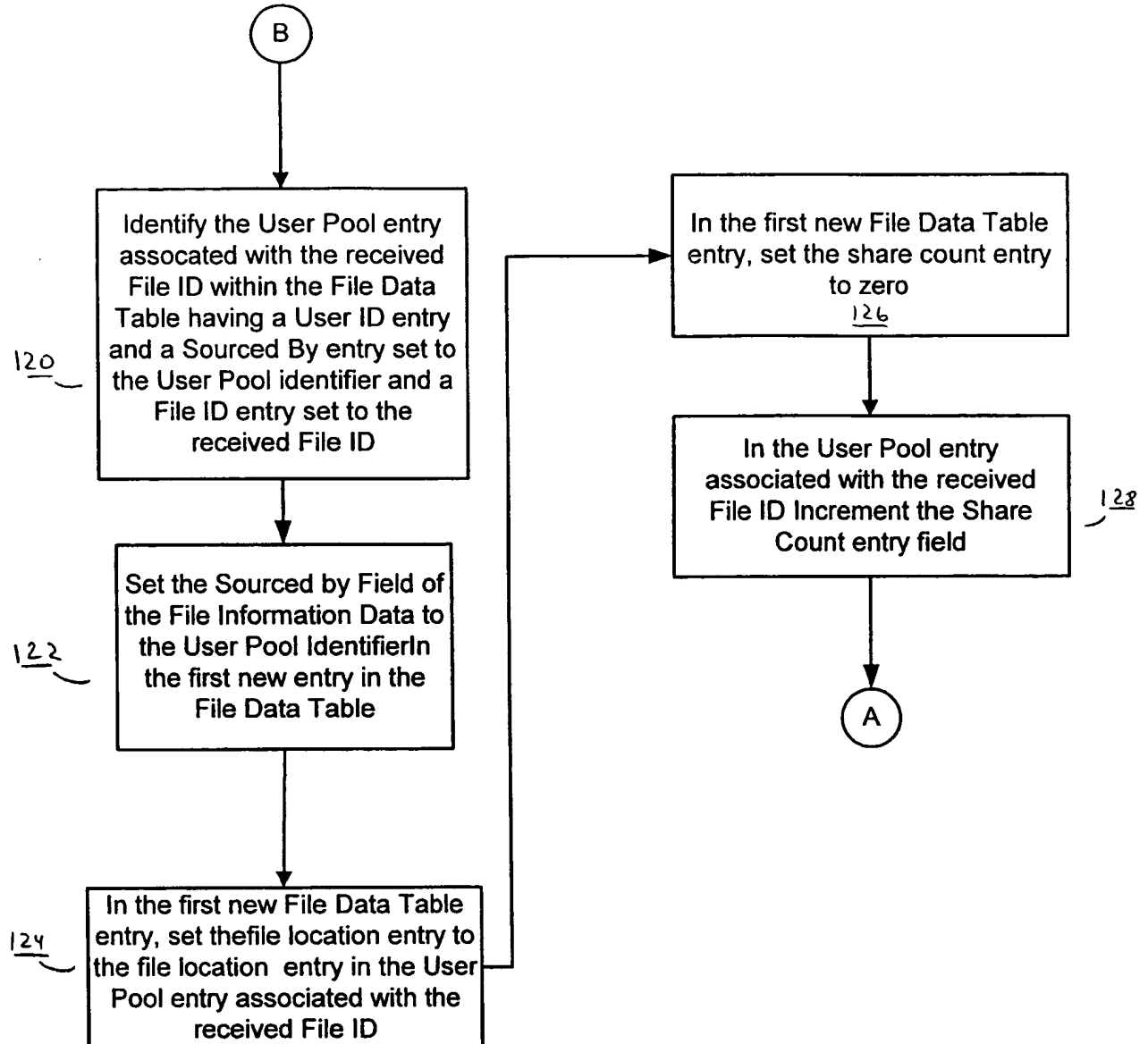


Fig. 1-B

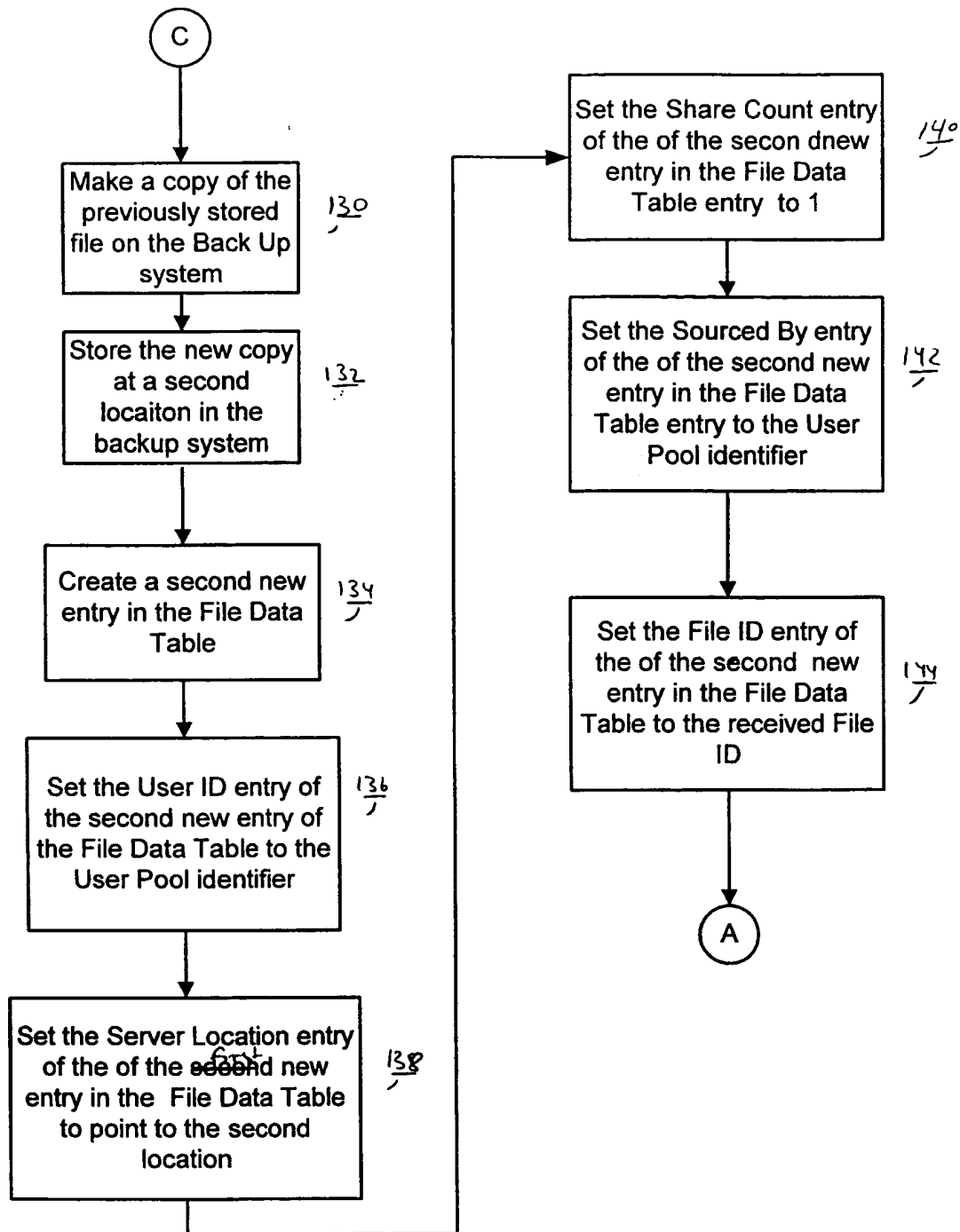


Fig. 1-C

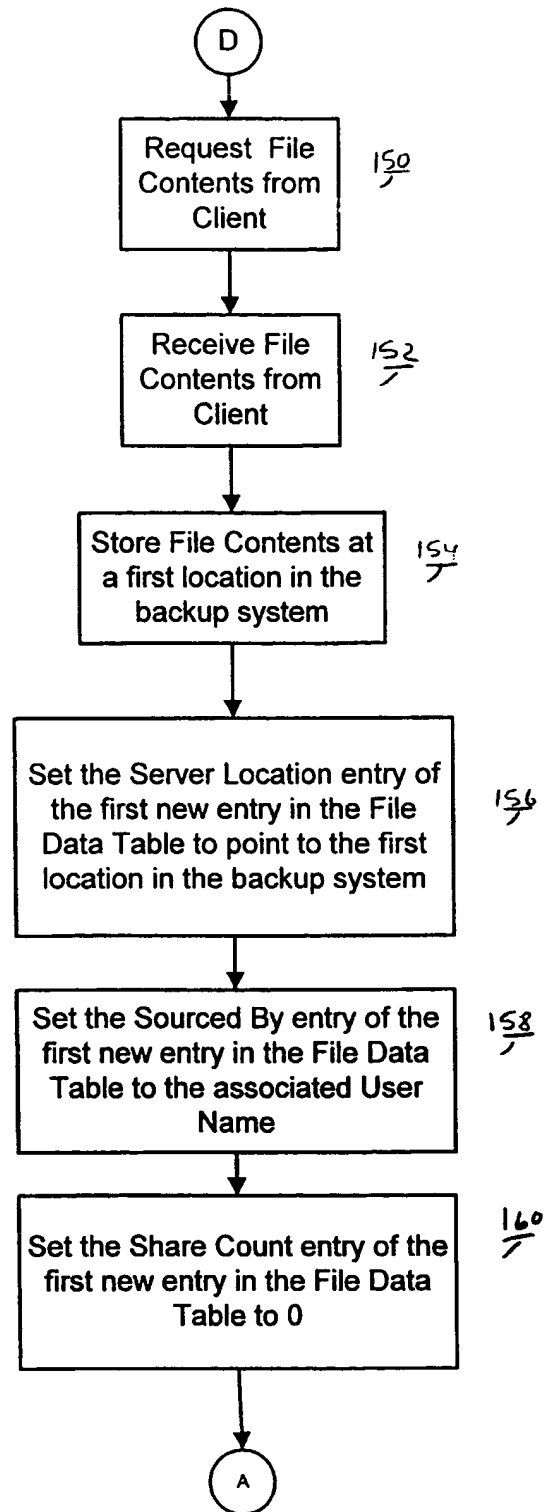


Fig 1-D

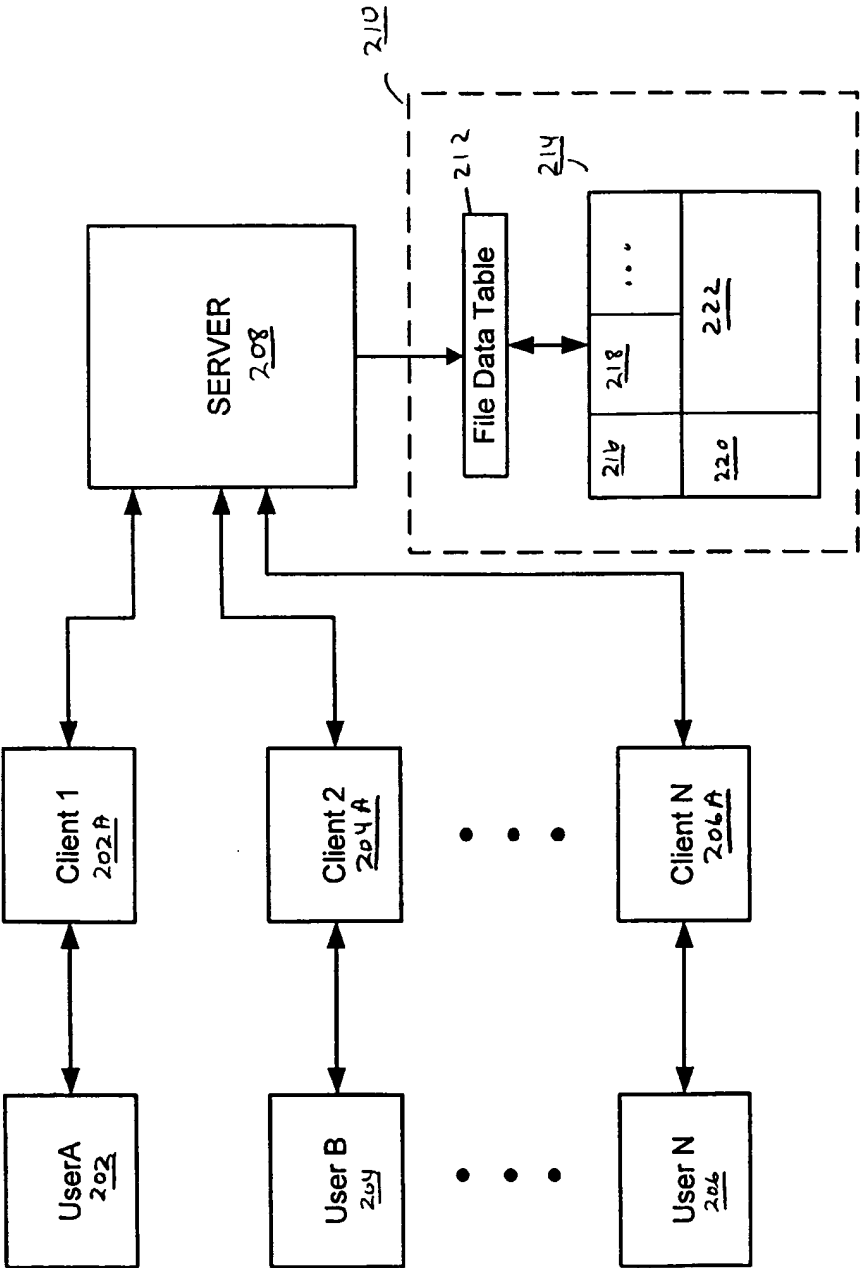


Fig. 2

212

302	304	306	308	310	312
User identifier	File identifier	File Location	Sourced by Identifier	Share Count	Other

Fig. 3

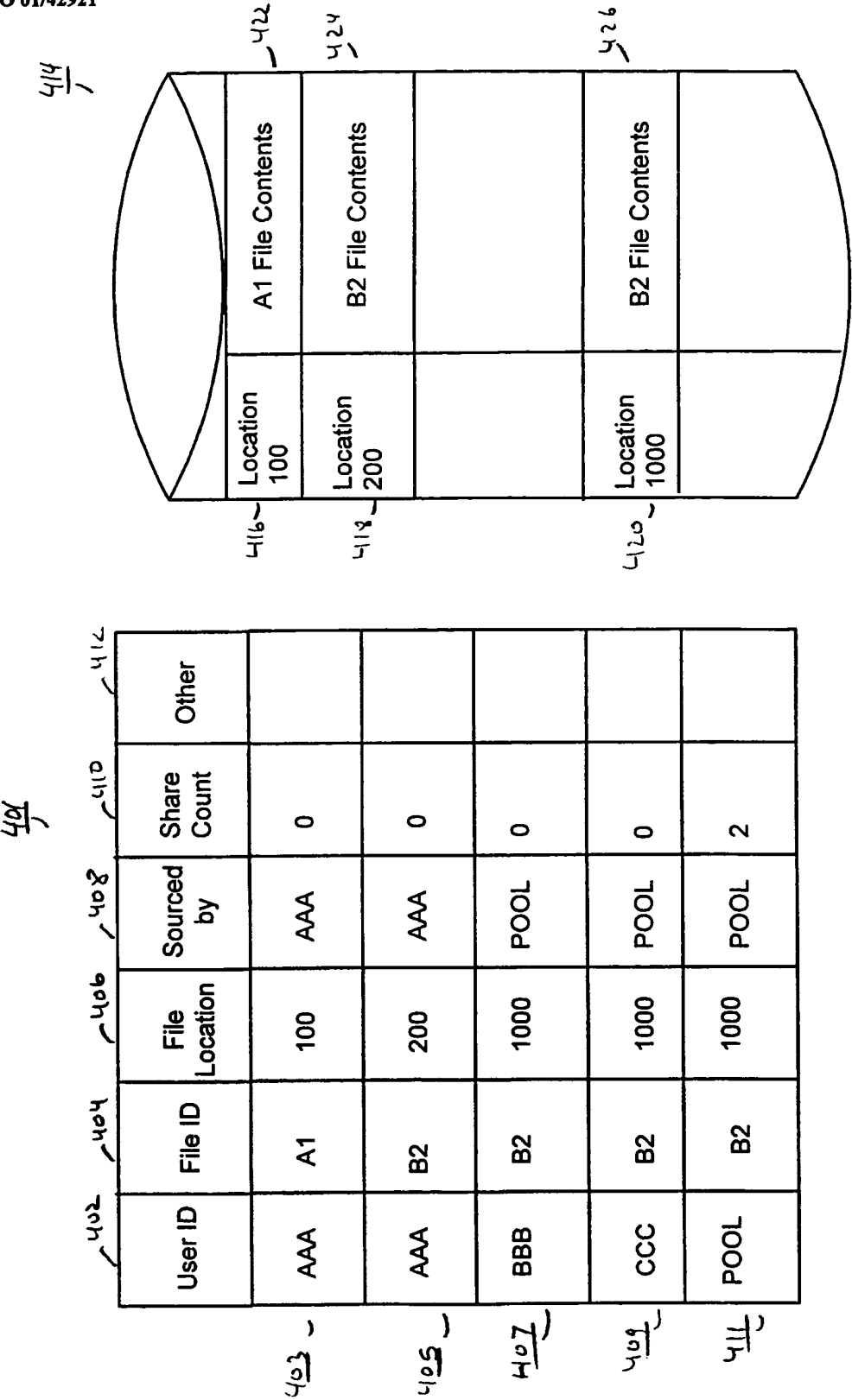


Fig. 4

INTERNATIONAL SEARCH REPORT

Int. national application No.

PCT/US00/42801

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 11/00

US CL : 714/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 714/6,20

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO, JPO, Derwent: backup, table, entry, archive, file**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,596,707 A (OHTA) 21 January 1997 (21.1.1997)	1-8
A,P	US 6,148,412 A (CANNON) 14 November 2000 (14.11.2000)	1-8
A	US 5,940,507 A (CANE) 17 August 1999 (17.9.1999)	1-8



Further documents are listed in the continuation of Box C.



See patent family annex.

Special categories of cited documents:	
* "A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"Z" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

18 April 2001 (18.04.2001)

Date of mailing of the international search report

09 MAY 2001

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Robert Beausoliel *James R. Matthews*
Telephone No. (703) 305-3900